UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/076,194	02/14/2002	Mingkun Li	US020037	3189

24737 7590 11/15/2006

PHILIPS INTELLECTUAL PROPERTY & STANDARDS
P.O. BOX 3001
BRIARCLIFF MANOR, NY 10510

EXAMINER	
PIERRE, MYRIAM	

ART UNIT	PAPER NUMBER
2626	

DATE MAILED: 11/15/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/076,194	LI ET AL.	
	Examiner	Art Unit	
	Myriam Pierre	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *24 August 2006*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,2 and 4-20* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-,2 and 4-20* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

Response to Arguments

1. Applicant's arguments filed 08/24/06 have been fully considered but they are not persuasive.

Applicant argues that Basu et al. (referred to as Basu) (6,219,640) fails to suggest or describe determining "a maximum correlation value among a plurality of correlation values between the plurality of object features and the plurality of audio features wherein said correlation values are determined as the sum elements in a subset of said audio features elected from the group consisting of two or more of the following: average energy, pitch, zero crossing, bandwidth, band central, roll off, low ratio, spectral flux, or 12 MFCC components. However, the examiner is relying on Nevenka (2003/0108334) for this limitation, as Nevenka et al. teach feature extraction from a list consisting of energy, pitch, and bandwidth (Fig. 2 "processor extracts feature audio streams" page 6 page 6 paragraph 65 lines 9-11 "audio parameters... energy, pitch, and bandwidth").

Applicant argues that Nevenka fails to describe creating AV vectors from a subset of audio and video features and determined the AV vector for the frame as that vector having a maximum correlation value wherein the correlation values are determined from a subset of the audio and video features. Examiner respectfully disagrees. Nevenka teach vector distances to measure the location and classification segmentation (subset) to correlate the time

intervals/pixels, frames and visual features, therefore, Nevenka do teach creating AV vectors from a subset of audio and video features and determined the AV vector for the frame as that vector having a maximum correlation value wherein the correlation values are determined from a subset of the audio and video features, page 5, paragraph 50, lines 1-7 and page 6 paragraph 65 lines 1-35.

Claim Rejections - 35 USC § 103

2. Claims 1, 2, 4, 5, 8, 11, and 16-17 are rejected under 35 U.S.C. 103(a) as being unpatentable over Basu et al. (6,219,640) in view of Nevenka (2003/0108334).

As per claim 1, Basu et al. teach an audio-visual processing data (col. 13, lines 55-58) comprising;

an object detection module capable of providing a plurality of object features from the video data (Fig. 4 elements 10, 20, and 24; and col. 6 lines 48-51; col. 10 lines 12-25);

an audio processor module capable of providing a plurality of audio features from the video data (col. 3 lines 53-59; col. 4 lines 58-67 and col. 8 lines 42-46);

a processor coupled to the object detection and the audio segmentation modules (col.13 lines 31-41; col. 11 lines 10-14; and col. 6 lines 33-48), arranged to determine a maximum correlation value among a plurality of correlation values between the plurality of object features and plurality of audio features (a level of correlation between the signals, col. 2, lines 35-36) wherein said correlation values are determined as the sum of the elements in a subset of said audio features (col. 9 lines 40-49, col. 10 lines 1-11; and col. 9 lines 15-34; maximum score is calculated from terms of the inner sum approach, selects the highest and second highest score,

from the top scores of the face identification process, the identification, which includes audio and video, of the speaker is known);

Basu et al. do not explicitly teach the audio features consisting of: two or more of the following: average energy, pitch, zero crossing, bandwidth, band central, roll off, low ratio, spectral flux, or 12 MFCC components.

However, Nevenka et al., teach feature extraction from a list consisting of energy, pitch, and bandwidth (Fig. 2 "processor extracts feature audio streams" page 6 page 6 paragraph 65 lines 9-11 "audio parameters... energy, pitch, and bandwidth").

Therefore, it would have been obvious for one of ordinary skill at the time of invention to combine Basu et al.'s audio and visual speaker recognition into the adaptive environment system of Nevenka et al., because Nevenka et al. teach that this would provide a system that passively records and identifies various events that occur in the home or office and can index the events using information, this way, a user can easily retrieve individual events and sub events using plain language commands or the processing system can determine whether an action is necessary in response to the identified event, page 2 paragraph 27 lines 17-24.

As per claim 2, which depends on claim 1, Basu et al. teach a processor arranged to determine whether an animated object in the video data is associated with audio (determine the level of correlation between the signals, col.2, lines 35-36).

As per claim 4, which depends on claim 2, Basu et al. teach that the animated object is a face (locate and track a face, other facial features, col 4, Lines 12-13) and where the processor is

arranged to determine whether the face is speaking (phonetic/visemic information from the geometry of the lip contour and its time dynamics, col. 10, Lines 53-55).

As per claim 5, which depends on claim 4, Basu et al. teach wherein the plurality of object features are eigenfaces that represent global features of the face (in "Distance from Face Space" DFFS, Lines, col 7. lines 32-35, feature vectors, col. 8, lines 7-8).

As per claims 8, 15 and 16, Basu et al. teach identifying a speaking person (speaker recognition and utterance verification, title) within video data, the method comprising:

- receiving video data including image (fig 1, element 4) and audio (figure 1, element 6) information,

- determining a plurality of face image features from one or more faces in the video data (sub-features, hairline, chin mouth, eyes, eyebrows, col 7, lines 55-57), determining a plurality of audio features related to audio information (extracts spectral features, col. 4, lines 61-63),

calculating correlation values between the plurality of face image features and the audio features (a level of correlation between the signals, lines 34-35), and

determining the speaking person based on a maximum of the correlation values (highest score identified as the speaker, col 10, lines 10-11; col. 9 lines 40-49, col. 10 lines 1-5, and col. 7 lines 6-25).

wherein said correlation values are determined as the sum of the elements of a subset between said audio features and selected object features (col. 9 lines 40-49, col. 10 lines 1-5, col. 7 lines 6-25, col. 11 lines 10-31, col. 8 lines 5-20, and col. 9 lines 15-34).

As per claim 11 and 17, which depend on claims 8 and 16, Basu et al. teach a determining step where it includes determining the speaking person based upon the one or more faces that has the largest correlation (highest combined score is identified as the speaker col 10, lines 10-11).

3. Claims 6-7 are rejected under 35 USC 103(a) as being unpatentable over Basu et al. (6,219,640) in view of Nevenka (2003/0108334) in further view of Bradford et al. (2002/0103799).

As per claim 6, which depends on claim 1, Neither Basu et al. nor Nevenka et al. explicitly teach a latent semantic indexing (LSI) module (coupled to the processor) that preprocesses the plurality of object features and the plurality of audio features before the correlation is performed.

However, Bradford teaches that to latent semantic indexing can be used to process both audio and text information vectors (para. 0079, lines 8-10).

Therefore, it would have been obvious for one of ordinary skill at the time of invention to combine the Audio-Visual speaker recognition into the adaptive environment of Basu et al. in view of Nevenka, into the audio and visual comparison of Bradford, because an artisan of ordinary skill in the art would want to provide a meaningful description of equivalents, (Bradford para. 0079).

As per claim 7, which depends on claim 6, Neither Basu et al. nor Nevenka et al.

explicitly teach a latent semantic indexing module including a singular value decomposition (SVD) module.

However, Bradford teaches using a SVD module (figure 2, para(0029)) to reduce term x Doc matrix to a product of three matrices.)

Therefore, it would have been obvious for one of ordinary skill at the time of invention to combine the Audio-Visual speaker recognition into the adaptive environment of Basu et al. in view of Nevenka, into the audio and visual comparison of Bradford, because an artisan of ordinary skill in the art would want to provide reduced matrix to a product of three matrices, (Bradford para 29).

4. Claim 9-10,12-14, and 18-20 are rejected under 35 USC 103(a) as being unpatentable over Basu et al. (6,219,640) in view of Nevenka et al. (2003/0108334), in further view of Wang et al.(Multimedia Content Analysis).

As per claim 9, which depends on claim 8, Neither Basu et al. nor Nevenka et al. explicitly teach normalizing the vectors containing the video/audio features.

However, Wang et al. teach normalizing these vectors (normalized correlation matrix pg 20, lines 2).

Therefore, it would have been further obvious to one having ordinary skill in the art at the time of invention to combine the Audio-Visual speaker recognition into the adaptive environment of Basu et al. in view of Nevenka, into the Multimedia Content Analysis of Wang et al, because an artisan of ordinary skill in the art would want to better interpret the correlation, if

any, that exists between the feature vectors, to see if they provide independent information, (Wang et al., p19, col 2, Lines 1-5).

As per Claims 10 and 18, which depend on claims 8 and 15, Neither Basu et al. nor Nevenka et al. explicitly teach performing a singular value decomposition on the normalized face image features and audio features.

However, Wang et al. do teach SVD on a normalized correlation matrix (pg 20, col 1, line 1 and col 2, Lines 4-5 (KLT-Karhunen Loeve transform).

Therefore, it would have been obvious for one of ordinary skill at the time of invention to combine the Audio-Visual speaker recognition into the adaptive environment of Basu et al. in view of Nevenka, into the Multimedia Content Analysis of Wang et al, because an artisan of ordinary skill in the art would want to decorrelate the features with KLT, (Wang et al. page 20 col. 2 para 1).

As per Claims 12 , Neither Basu et al. nor Nevenka et al. explicitly teach a calculating step which includes forming a matrix of the face image features and the audio features.

However, Wang et al. do teach combining the two in a single matrix (14 audio features, last six motion features, figure 9 and pg 20, Lines 8-10).

It would have been further obvious to one having skill in the art at the time of invention to combine the Audio-Visual speaker recognition into the adaptive environment of Basu et al. in view of Nevenka, into the Multimedia Content Analysis of Wang et al, because an artisan of

ordinary skill in the art would want the dependence among features within the same and across different modalities to be computed (Wang et al. pg 19, lines 5-8).

As per Claims 13 and 19, which depends on claims 12 and 18, Neither Basu et al. nor Nevenka et al. explicitly teach performing an optimal approximate fit using smaller matrices as compared to full rank matrices formed by the face image features and audio features.

However, Wang et al. do teach using SVD to allow for dimensionality reduction (pg 10, Lines 18-19).

It would have been further obvious to one having skill in the art at the time of invention to combine the Audio-Visual speaker recognition into the adaptive environment of Basu et al. in view of Nevenka, into the Multimedia Content Analysis of Wang et al, because an artisan of ordinary skill in the art would want to decorrelate the features, (Wang et al. page 20 col. 2 para 1).

As per claims 14 and 20, which depends on claims 13 and 19, Basu et al. teach

choosing the rank of the smaller matrices to remove noise and unrelated information from the full rank matrices (col. 14 lines 31-59).

Conclusion

5. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Myriam Pierre whose telephone number is 571-272-7611. The examiner can normally be reached on Monday - Friday from 5:30 a.m. - 2:00p.m.

6. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

7. Information as to the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

11/11/06

ABUL AZAD
PRIMARY EXAMINER

Myriam Pierre
AU 2626
11/07/06